Graph Limits and Parameter Testing

Christian Borgs
borgs@microsoft.com

Jennifer Chayes
jchayes@microsoft.com

László Lovász
lovasz@microsoft.com

Microsoft Research, Redmond, WA 98052

Vera T. Sós*
Rényi Institute
Budapest, Hungary
sos@renyi.hu

Balázs Szegedy
Institute for Advanced Study
Princeton, NJ
szegedyb@gmail.com

Katalin Vesztergombi
Dept. Comp. Sci.
Eötvös Loránd University
Budapest, Hungary
veszterk@cs.elte.hu

ABSTRACT

We define a distance of two graphs that reflects the closeness of both local and global properties. We also define convergence of a sequence of graphs, and show that a graph sequence is convergent if and only if it is Cauchy in this distance. Every convergent graph sequence has a limit in the form of a symmetric measurable function in two variables. We use these notions of distance and graph limits to give a general theory for parameter testing. As examples, we provide short proofs of the testability of MaxCut and the recent result of Alon and Shapira about the testability of hereditary graph properties.

Categories and Subject Descriptors

G.2.2 [**Mathematics of Computing**]: Discrete mathematics—*Graph theory*

General Terms

Theory

Keywords

Graph homomorphism, distance of graphs, convergence of graphs, graph limit, property testing

*Research supported by OTKA grants T032236, T038210, T042750

STOC'06, May 21–23, 2006, Seattle, Washington, USA.

1. INTRODUCTION

Imagine that we have a huge graph G, so large that we cannot describe it completely in any way. All we can do is sample a bounded number of nodes of G and look at the subgraph induced by them. What can we learn about G?

There are two related, but slightly different ways of asking this question, *property testing* and *parameter testing*. Our main concern here will be parameter testing; we'll show one way to obtain results about property testing from this, by considering the "edit distance" from the property as a parameter.

Parameter testing is easier to state. We may want to determine some parameter of G, say what is the edge density? Or how large is the density of the maximum cut? Of course, we'll not be able to determine the exact value of this parameter; the best we can hope for is that if we take a sufficiently large sample, we can find the approximate value of the parameter with large probability.

To be precise, a graph parameter f is *testable*, if for every $\varepsilon > 0$ there is a positive integer k such that if G is a graph with at least k nodes and we select a set X of k independent uniform random nodes of G, then from the subgraph $G[X]$ induced by them we can compute an estimate $\widetilde{f}(G[X])$ of f such that

$$\mathsf{P}(|f(G) - \widetilde{f}(G[X])| > \varepsilon) < \varepsilon.$$

It is an easy observation that we can always use $\widetilde{f}(G[X]) = f(G[X])$ (cf. [14]).

As a basic example, consider the density of maximum cuts (i.e., the number $\mathsf{maxcut}(G) = \mathsf{MaxCut}(G)/|V(G)|^2$, where $\mathsf{MaxCut}(G)$ is the size of the maximum cut in G). One of the first substantial results on property testing [13, 7] is that this parameter is testable. It is relatively easy to see (using high concentration results like Azuma's inequality) that if S is a sufficiently large random subset of nodes of G, then $\mathsf{maxcut}(G[S]) \geq \mathsf{maxcut}(G) - \varepsilon$: a large cut in G, when restricted to S, gives a large cut in $G[S]$. It is much harder, and in fact quite surprising, that if most subgraphs $G[S]$ have a large cut, then so does G.

Instead of estimating a numerical parameter, we may want to determine some property of G: Is G 3-colorable? Is it connected? Does it have a triangle? The answer will of course have some uncertainty. A precise definition was given by Rubinfeld and Sudan [18] and by Goldreich, Goldwasser and Ron [13], who also proved several fundamental results about this problem. (In the slightly different context of "additive approximation", closely related problems were studied by Arora, Karger and Karpinski [7].) The standard definition is as follows. A property $\mathcal{P}$ is said to be testable if, for every $\varepsilon > 0$, there is a positive integer k such that for all graphs G on at least k nodes, the following holds: If X is a k-node random subset of G, then, with high probability, the induced subgraph $G[X]$ has the property $\mathcal{P}$ whenever G has the property, and $G[X]$ does not have the property whenever G is ε-far from having the property. Here we say that a graph G on n nodes is ε-far from having the property $\mathcal{P}$ if all graphs G' which differ from G on at most εn^2 edges do not have the property $\mathcal{P}$.

Graph property testing has a large literature; see e.g. [11] for a survey. Many extensions deal with situations where we are allowed to sample more than a constant number of nodes of the large graph G; our concern will be the original setup, where the sample size is bounded. In this direction, a surprisingly general result was proved very recently by Alon and Shapira [5, 6]: they showed that every hereditary graph property is testable; more generally, the "edit distance" to a hereditary property is a testable parameter.

We develop a general theory of parameter testing, based on the notions of convergent graph sequences (introduced by Lovász and Szegedy [15]) and graph distances (introduced by Borgs, Chayes, Lovász, Sós and Vesztergombi [9]).

In Section 3, we introduce a notion of "distance" of two graphs that is suitable for the study of parameter testing. In Section 4, we define the notion of convergence of graphs and characterizations of the limit object. Informally, convergent sequences will be just those sequences that are Cauchy in this metric, and testable parameters will be just those graph parameters that are continuous in this metric.

As an illustration of our results, we show how the testability of the density of maximum cut (and extensions to multiway cuts) follows from our results, and sketch a reasonably simple way to derive the recent result of Alon and Shapira mentioned above.

2. PRELIMINARIES: GRAPH HOMOMORPHISMS

A *graph parameter* is a function defined on simple graphs that is invariant under isomorphism.

For two graphs G and H, a homomorphism from G to H is an adjacency preserving map $V(G) \to V(H)$. When G is large and H is small, a homomorphism from G to H is often called an *H-coloring* of G. Let $\hom(G, H)$ denote the number of homomorphisms from G to H.

Let G be a graph on n nodes, and F be a graph on k nodes. We define $t(F, G)$ to be the probability that a random map of $V(F)$ into $V(G)$ is a homomorphism, i.e.,

$$t(F, G) = \frac{\hom(F, G)}{n^k}.$$

We call $t(F, G)$ the *homomorphism density* (of F in G).

Sometimes it is more convenient to consider the number of injective homomorphisms $\mathrm{inj}(F, G)$, and the number of embeddings as induced subgraphs, $\mathrm{ind}(F, G)$. The corresponding densities are defined by

$$t_{\mathrm{inj}}(F, G) = \frac{\mathrm{inj}(F, G)}{n(n-1)\dots(n-k+1)}$$

and

$$t_{\mathrm{ind}}(F, G) = \frac{\mathrm{ind}(F, G)}{n(n-1)\dots(n-k+1)}.$$

These quantities don't carry much new information: if G is large enough then $t_{\mathrm{inj}}(F, G) \approx t(F, G)$, and $t_{\mathrm{ind}}(F, G)$ can be expressed in terms of $t_{\mathrm{inj}}(F, G)$ by inclusion-exclusion.

We extend these notions to the case when the target graph G is weighted. A *weighted graph* G is a graph with a weight $\alpha_i(G)$ associated with each node and a weight $\beta_{ij}(G)$ associated with each edge ij. In this paper we assume that $\alpha_i(G) > 0$ and $0 \le \beta_{ij}(G) \le 1$. An edge with weight 0 will play the same role as no edge between those nodes. The adjacency matrix of a weighted graph is obtained by replacing the 1's in the adjacency matrix by the weights of the edges. Let $\alpha_G = \sum_{i \in V(G)} \alpha_i(G)$ denote the total nodeweight of G. An *unweighted graph* is a weighted graph where all the node- and edgeweights are 1.

The notion of homomorphisms can be easily extended to the case when the target graph G is weighted:

$$\hom(F, G) = \sum_{\phi:\ V(F) \to V(G)} \prod_{u \in V(F)} \alpha_{\phi(u)}(G) \times \prod_{uv \in E(F)} \beta_{\phi(u),\phi(v)}(G)$$

where the sum runs over all maps from $V(F)$ to $V(G)$. The homomorphism density is now defined as

$$t(F, G) = \frac{\hom(F, G)}{{\alpha_G}^{|V(F)|}}.$$

Example 1 If G is a simple graph, then $\hom(K_3, G)$ is 6 times the number of triangles in G. So $t(K_3, G)$ is the triangle density.

Example 2 If F is a simple graph, then $\hom(F, K_q)$ is the number of q-colorations of the graph G.

Example 3 Let G consist of two nodes, connected by an edge, and with a loop at one of the nodes. Then $\hom(F, G)$ is the number of independent sets of nodes in F.

Example 4 Let G consist of two nodes, connected by an edge with weight 2, and with a loop of weight 1 at each of the nodes. Then

$$|\mathsf{MaxCut}(F) - \log_2 \hom(F, G)| = O(|V(F)|),$$

so except for sparse graphs, $\log_2 \hom(F, G)$ gives a very good approximation for the maximum cut in G.

3. DISTANCES OF GRAPHS

3.1 Definition of graph distances

We want to define a measure of similarity of two large graphs. There is more than one very reasonable definition; but one of these, the "rectangle" or "cut" distance, will be particularly useful.

We have to proceed in two steps. It is easier to define the distance when the two graphs have the same set of nodes; but then we have to deal with finding the optimal overlay. It turns out that it is worth extending the arguments to "fractional overlays" (in the spirit of fractional solutions to integer programs, often used in combinatorial optimization).

3.1.1 Labeled graphs

Let G and G' be two unweighted labeled graphs with the same set of n of nodes. We want to define a notion of distance between them that reflects structural similarity. A natural distance measure is the "edit distance":

$$d_{\text{edit}}(G,G') = \frac{|E(G) \triangle E(G')|}{n^2},$$

where for convenience we divided by n^2, so that the distance of two graphs is always between 0 and 1.

This distance notion is, however, often too restrictive: For example, the distance of two random graphs with the same density is of constant order (with large probability), even though two random graphs are structurally very similar. For our purposes, the following distance function between unweighted labeled graphs will be more useful:

$$d_{\square}(G,G') = \frac{1}{n^2} \max_{S,T \subseteq \{1,\dots,n\}} |e_G(S,T) - e_{G'}(S,T)|,$$

where $e_G(S,T)$ is the number of edges joining S and T (with the edges in $S \cap T$ counted twice, so that $e_G(S,S)$ is twice the number of edges in G). Note that we are dividing by n^2 and not by $|S| \cdot |T|$, so the contribution of a pair S,T is at most $|S| \cdot |T|/n^2$. Thus small sets of size $o(n)$ play no role when measuring the distance.

If G is weighted, let

$$e_G(S,T) = \sum_{i \in S} \sum_{j \in T} \alpha_i(G) \beta_{ij}(G) \alpha_j(G).$$

For the case when G and G' have the same nodeweights, we define

$$d_{\square}(G,G') = \frac{1}{\alpha_G^2} \max_{S,T \subseteq \{1,\dots,n\}} |e_G(S,T) - e_{G'}(S,T)|.$$

There are several versions of the $d_{\square}$ distance that differ from each other only by absolute constant factors. For example, we could restrict the maximization in the definition to $S = T$. Another notable variation is obtained by considering an appropriate semidefinite relaxation of the maximization problem involved in the definition. The fact that this only changes by a factor of at most $2.78\dots$ is equivalent to Grothendieck's inequality in functional analysis. This shows that $d_{\square}(G,G')$ can be computed in polynomial time up to an absolute constant factor (see [4]).

3.1.2 Unlabeled graphs with the same number of nodes

Now assume that G and G' are unlabeled graphs on n nodes with nodeweights 1. We then define

$$\widehat{\delta}_{\square}(G,G') = \min_{\tilde{G},\tilde{G}'} d_{\square}(\tilde{G},\tilde{G}'),$$

where $\tilde{G}$ and $\tilde{G}'$ range over all labelings of G and G', respectively.

Consider any labeling that attains the minimum in this definition, and identify the nodes of G and G' with the same label. In this case, we say that G and G' are *optimally overlaid.*

3.1.3 Unlabeled graphs with different number of nodes

To define the distance of two unlabeled graphs with different number of nodes, say G with n nodes and G' with n' nodes, a first idea is to blow up each node of G into n' twins, and each node of G' into n twins, so that both graphs now will have nn' nodes. Here, as usual, two nodes in a graph are called twins if they have the same neighborhood. An improved version of this idea is to match up the nodes "fractionally".

It will be convenient to consider weighted graphs G and G' right away, and assume that the sum of nodeweights is 1 (just scale the nodeweights of each graph). Let X be a nonnegative $n \times n'$ matrix such that

$$\sum_{u=1}^{n'} X_{iu} = \alpha_i \quad \text{and} \quad \sum_{i=1}^{n} X_{iu} = \alpha'_u.$$

We think of X_{iu} as the portion of node i that is mapped onto node u. We call such a matrix X a *fractional overlay* of G and G', and use $\mathcal{X}(G,G')$ to denote the set of all these fractional overlays. If we view $\alpha(G)$ and $\alpha(G')$ as probability distributions on $V(G)$, then every $X \in \mathcal{X}(G,G')$ is a *coupling* of these distributions.

For each fractional overlay, we construct the following two weighted graphs, $G[X]$ and $G'[X^\top]$. The nodes of $G[X]$ are all pairs (i,u) where $1 \le i \le n$ and $1 \le u \le n'$. The weight of the node (i,u) is X_{iu}, the weight of the edge $((i,u),(j,v))$ is β_{ij}. The other graph $G'[X^\top]$ is defined similarly, except that the roles of i and u are interchanged.

Now the node sets of $G[X]$ and $G'[X^\top]$ are labeled by the same set of pairs (i,u), and they have the same nodeweights, so their $d_{\square}$ distance is well defined. Minimizing over all fractional overlays, we therefore obtain a well-defined distance between any two weighted unlabeled graphs G and G' each with total nodeweight 1:

$$\delta_{\square}(G,G') = \min_{X \in \mathcal{X}(G,G')} d_{\square}(G[X], G'[X^\top]).$$

This distance can be expressed in terms of the original graphs G and G' according to

$$\delta_{\square}(G,G') = \min_{X \in \mathcal{X}(G,G')} \max_{S,T \subseteq V \times V'} \Big| \sum_{\substack{(i,u) \in S \\ (j,v) \in T}} X_{iu} X_{jv} (\beta_{ij}(G) - \beta_{uv}(G')) \Big|.$$

If G or G' do not have total nodeweight 1, we define their distance as the distance of the rescaled graphs $\tilde{G}$ and $\tilde{G}'$ obtained from G and G' by dividing the nodeweights of G and G' by α_G and $\alpha_{G'}$, respectively.

Of course, the above definition also applies if G and G' are simple graphs with the same number of nodes, but it may give a different value than discussed above in section 3.1.2. It is trivial that $\delta_{\square}(G,G') \le \widehat{\delta}_{\square}(G,G')$. An inequality in the other direction also holds [9], so that

$$\delta_{\square}(G,G') \le \widehat{\delta}_{\square}(G,G') \le C \delta_{\square}(G,G')^{1/4} \tag{1}$$

with an absolute constant C. We believe that a much stronger bound holds, probably

$$\delta_{\square}(G,G') = \Omega(\widehat{\delta}_{\square}(G,G')).$$

This notion of graph distance allows us to give a very simple formulation of Szemerédi's Regularity Lemma, at least in its weaker (but more effective) form given by Frieze and Kannan [12]. For a graph G and partition $\mathcal{P} = (V_1, \dots, V_k)$ of $V(G)$, define the weighted complete graph $G_{\mathcal{P}}$ for which $V(G_{\mathcal{P}}) = V(G)$ and every $u \in V_i$ and $v \in V_j$ is connected by an edge with weight $e_G(V_i, V_j)/(|V_i| \cdot |V_j|)$.

Lemma 3.1 (Frieze-Kannan) *For every graph G and every $\varepsilon > 0$, there exists a partition of $V(G)$ with at most $2^{2/\varepsilon^2}$ classes such that $d_\square(G, G_{\mathcal{P}}) \le \varepsilon$.*

(At the cost of increasing the bound on $|\mathcal{P}|$, we could assume that the classes of $\mathcal{P}$ have equal size.)

Corollary 3.2 *For every graph G and every $\varepsilon > 0$, there exists a weighted graph H with at most $2^{2/\varepsilon^2}$ nodes such that $\delta_\square(G, H) \le \varepsilon$.*

The following inequality was proved in [15]; it is also closely related to the "Counting Lemma" in the theory of Szemerédi partitions:

Lemma 3.3 *For any pair of weighted graphs G, G' and any simple graph F,*

$$|t(F, G) - t(F, G')| \le |E(F)| \cdot \delta_\square(G, G').$$

It will be more convenient to prove this inequality after we introduce the "limit objects" in Section 4.

3.2 The distance of a sample

The following lemma is a consequence of a more general result in [2]; see also [9] for a simpler proof of a weaker version (which would still be enough for us).

Lemma 3.4 *Let G_1 and G_2 be two weighted graphs on the same set of nodes V with nodeweights 1, and let $\varepsilon = d_\square(G_1, G_2)$. Let $\delta > 0$, $k \ge 10^{10} \log(2/\varepsilon)/(\varepsilon^4\delta^5)$, and let S be a random k-subset of V. Then with probability at least $1 - \delta$,*

$$2^{-10}\varepsilon \le d_\square(G_1[S], G_2[S]) \le 10^7 \varepsilon/\sqrt{\delta}.$$

Now we come to the main theorem about sampling.

Theorem 3.5 *Let G be a simple graph, let $\varepsilon, \delta > 0$ and $k \ge 2^{\Theta(1/(\delta\varepsilon^2))}$. Let S be a random k-subset of V. Then with probability at least $1 - \delta$,*

$$\delta_\square(G, G[S]) \le \varepsilon.$$

Proof. We use the weak Regularity Lemma 3.1. Let $\mathcal{P}$ be a partition of $V(G)$ with $q = 2^{2/(\varepsilon\sqrt{\delta}10^{-8})^2}$ classes so that $d_\square(G, G_{\mathcal{P}}) \le \varepsilon\sqrt{\delta}10^{-8}$. Consider a random subset S of $V(G)$ with k elements. Then

$$\begin{aligned}\delta_\square(G, G[S]) \le &\delta_\square(G, G_{\mathcal{P}}) + \delta_\square(G_{\mathcal{P}}, G_{\mathcal{P}}[S]) \\ &+ \delta_\square(G_{\mathcal{P}}[S], G[S]). \qquad (2)\end{aligned}$$

On the right hand side, the first term is at most $\varepsilon\sqrt{\delta}10^{-8} \le \varepsilon 10^{-8}$ by the the choice of $\mathcal{P}$. To estimate the second term, notice that both $G_{\mathcal{P}}$ and $G_{\mathcal{P}}[S]$ are obtained by blowing up the same weighted graph H with q nodes (we may assume that $\alpha_H = 1$): in $G_{\mathcal{P}}$, we blow up each node $i \in V(H)$ into a set V_i with $|V_i| = \alpha_i(H)|V(G)|$; in $G_{\mathcal{P}}[S]$, we blow up each node into $|V_i \cap S|$ nodes. Since S is a random sample, with large probability all these sizes will be close to their expectation, which in turn is closed to $\alpha_i k$ provided k/n is small. Blowing up each node of $G_{\mathcal{P}}[S]$ into $|V(G)|/k$ copies, we get a graph H' that can also be obtained from H by blowing up each node i into approximately $\alpha_i(H)|V(G)|$ copies. This argument can be made precise by choosing appropriate fractional overlays, leading to the estimate

$$\delta_\square(G_{\mathcal{P}}, G_{\mathcal{P}}[S]) = \delta_\square(G_{\mathcal{P}}, H') \le 2\sum_{i=1}^{q}\Big|\alpha_i - \frac{1}{k}|S \cap V_i|\Big|.$$

Using standard second moment bounds, the right hand side can be made smaller than $\varepsilon/3$ with probability at least $1 - \delta/2$ provided $k = O(\epsilon n)$ and $q = O(\delta\varepsilon^2 k)$.

Finally, to see that the third term in (2) is at most $\varepsilon/3$, we invoke Lemma 3.4: this says that the samples $G[S]$ and $G_{\mathcal{P}}[S]$ are close in the $d_\square$ metric, completing the proof for $k = O(\epsilon n)$. The case $k \ge \varepsilon n$ is much easier and left to the reader, see also [9]. □

Corollary 3.6 *For every simple graph G and $\varepsilon > 0$ there is simple graph H with at most $2^{\Theta(1/\varepsilon^2)}$ nodes such that*

$$\delta_\square(G, H) < \varepsilon.$$

This corollary can be thought of as a strengthening of the (weak) Szemerédi lemma in two directions. First, it says that the approximating weighted graph can have 0-1 weights; second, that it can be obtained just by drawing a random sample. The theorem also explains why parameter testing works: if a parameter is "continuous" in the $\delta_\square$ metric, then with large probability, it will not change much if G is replaced by the random sample $G[S]$.

In terms of metric spaces, Corollary 3.6 says that the set of simple graphs with at most $2^{c/\varepsilon^2}$ nodes form an ε-net in the metric space of all simple graphs with the metric $\delta_\square$.

4. CONVERGENT GRAPH SEQUENCES AND THEIR LIMITS

4.1 Definitions

Let (G_n) be a sequence of unweighted simple graphs. Throughout this paper we assume that $|V(G_n)| \to \infty$. We say that this sequence is *convergent*, if the sequence $t(F, G_n)$ has a limit for every simple graph F. Note that it would be enough to assume this for connected graphs F; indeed, if $F_1, \dots, F_k$ are the connected components of F, then $t(F, G_n) = \prod_i t(F_i, G_n)$, and so if each factor here is convergent, so is $t(F, G_n)$. Furthermore, it is not hard to see that instead of $t(F, G_n)$, we could stipulate the convergence of $t_{\mathrm{inj}}(F, G_n)$ or $t_{\mathrm{ind}}(F, G_n)$.

This notion of convergence corresponds to the notion of distance we have introduced [9]:

Theorem 4.1 *A graph sequence (G_n) is convergent if and only if it is Cauchy in the $\delta_\square$ metric.*

A "finite" implication of this theorem is that if two huge graphs on the same number of nodes have approximately the same subgraph densities, then they can be overlaid so that they will be close in the $d_\square$ distance. Note that this statement provides a certain converse to Lemma 3.3.

PROOF. The "if" part of this result follows immediately from Lemma 3.3. To prove the "only if" part, consider a convergent graph sequence (G_n). It is easy to see that this implies that $t_{\rm ind}(F, G_n)$ has a limit for every F, which is equivalent to saying that for every fixed k, the distribution of $G_n[S_n]$ (where S_n is a random k-subset of $V(G_n)$) has a limit, so there is an n_0 such that for $n, m \geq n_0$, the total variation distance of the distributions of $G_n[S_n]$ and $G_m[S_m]$ is less than $\varepsilon/3$, implying that we can couple them in such a way that $\delta_\square(G_n[S_n], G_m[S_m]) = 0$ with probability at least $1 - \varepsilon/3$. By Theorem 3.5, if we choose k large enough, we have that for each n

$$\delta_\square(G_n, G_n[S_n]) \leq \frac{\varepsilon}{6},$$

and each m

$$\delta_\square(G_m, G_m[S_m]) \leq \frac{\varepsilon}{6}$$

with probability at least $1-\varepsilon/3$, and so with probability at least $1-2\varepsilon/3$, we have that $\delta_\square(G_n, G_m) \leq \delta_\square(G_n, G_n[S_n]) + \delta_\square(G_m, G_m[S_m]) \leq \varepsilon/3$. □

One can define a "limit object" for every convergent graph sequence; in fact, there are several quite different and useful descriptions of the limit object [15]. In this extended abstract, we are going to use the following. Let $\mathcal{W}$ denote the set of all bounded measurable functions $W : [0,1]^2 \to \mathbb{R}$ such that $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$, and let $\mathcal{W}_0 = \{W \in \mathcal{W} : 0 \leq W \leq 1\}$. Let $W \in \mathcal{W}_0$, and let F be a simple graph with $V(F) = \{1, \dots, k\}$. We define

$$t(F, W) = \int_{[0,1]^k} \prod_{ij \in E(F)} W(x_i, x_j)\, dx.$$

It is easy to see that for every weighted graph H with $\alpha_H = 1$, the graph parameter $t(., H)$ is a special case: we consider the adjacency matrix of H, and replace each entry (i, j) by a square of size $\alpha_i(H) \times \alpha_j(H)$ with the constant function β_{ij} on this square. The function W_H obtained this way satisfies $W_H \in \mathcal{W}_0$, and

$$t(F, H) = t(F, W_H)$$

for every simple graph F.

It was proved in [15] that

Proposition 4.2 *For every convergent graph sequence (G_n) there is a function $W \in \mathcal{W}_0$ such that*

$$\lim_{n\to\infty} t(F, G_n) = t(F, W)$$

for every finite simple graph F.

We call this function W the *limit* of the sequence (G_n). Note that the limit is not uniquely determined. For example, the function $W(1-x, 1-y)$ also satisfies the conditions. We'll return to this question later.

4.2 Examples

Example 5 (Random graphs) Let $\mathbf{G}(n, p)$ be a random graph on n nodes with edge-density $0 \leq p \leq 1$; then it is not hard to prove (using high concentration results) that the sequence $(\mathbf{G}(n, p), n = 1, 2, \dots)$ is convergent with probability 1. In fact, $t(F, \mathbf{G}(n, p))$ converges to $p^{|E(F)|}$ with probability 1, and so (with probability 1) the limit of $\mathbf{G}(n, p)$ is the constant function $W = p$.

Example 6 (Quasirandom graphs) A graph sequence is quasirandom with density p if and only if it converges to the constant function p (cf. [10] for the definition and various characterizations of quasirandom graph sequences).

Example 7 (Half-graphs) Let $H_{n,n}$ denote the bipartite graph on $2n$ nodes $\{1, \dots, n, 1', \dots, n'\}$, where i is connected to j' if and only if $i \leq j$. It is easy to see that this sequence is convergent, and its limit is the function

$$W(x, y) = \begin{cases} 1, & \text{if } |x - y| \geq 1/2, \\ 0, & \text{otherwise.} \end{cases}$$

Example 8 (Uniform attachment) Various sequences of growing graphs, motivated by (but different from) internet models, are also convergent. We define a *uniform attachment graph sequence*: if we have a current graph G_n with n nodes, then we create a new isolated node, and then for every pair of previously nonadjacent nodes, we connect then with probability $1/n$.

One can prove that with probability 1, the sequence (G_n) has a limit, which is the function $W(x, y) = \min(x, y)$. From this, it is easy to calculate that with probability 1, the edge-density of G_n tends to $\int W = 1/3$. More generally, the density of copies of any fixed graph F in $G(n)$ tends (with probability 1) to $t(F, W)$, which can be evaluated by a simple integration.

4.3 Some properties of limit functions

4.3.1 W-random graphs

Every function in $\mathcal{W}_0$ arises as a limit. One proof of this fact is through the construction of certain random graphs that are of interest in their own right. Given a function $W \in \mathcal{W}_0$ and an integer $n > 0$, we can generate a *W-random graph* $\mathbf{G}(n, W)$ on nodes $\{1, \dots, n\}$ as follows: We generate n independent samples $X_1, \dots, X_n$ from the uniform distribution on $[0, 1]$, and for all $i, j \in \{1, \dots, n\}$, we connect i and j by an edge ij with probability $W(X_i, X_j)$ (making an independent decision for every pair). Using Azuma's inequality, one can prove [15]:

Theorem 4.3 *The graph sequence $\mathbf{G}(n, W)$ is convergent with probability 1, and its limit is the function W.*

4.3.2 Distances of functions

An analogue of the $d_\square$ distance can be defined for functions $U, W \in \mathcal{W}_0$; in fact, it can be defined as a norm:

$$\|U\|_\square = \sup_{S,T\subseteq[0,1]} \Big| \int_{S\times T} U(x, y)\, dx\, dy \Big|$$

(the distance of two functions U, W is $\|U - W\|_\square$). Recall that the standard L_1 norm is defined as

$$\|U\|_1 = \int_{[0,1]^2} |U(x, y)|\, dx\, dy$$

so that

$$\|U\|_\square \leq \|U\|_1. \tag{3}$$

It is not hard to see that this norm could also be defined by

$$\|U\|_\square = \sup_{f,g} \Big| \int_{[0,1]^2} f(x)g(y)U(x, y)\, dx\, dy \Big|,$$

where $f, g : [0,1]^2 \to [0, 1]$ are integrable functions.

Analogously to graphs, we define a version which corresponds to the best "overlaying". We define a *measure preserving bijection* as a bijective map $\phi : [0,1] \mapsto [0,1]$ such that both ϕ and ϕ^{-1} are measurable and $\lambda(\phi(U)) = \lambda(U)$ for every measurable set $U \subseteq [0,1]$ (where λ is the Lebesgue measure). For $W \in \mathcal{W}_0$, we define W^ϕ by $W^\phi(x,y) = W(\phi(x),\phi(y))$. With this notation, let

$$\delta_\square(U,W) = \inf_\phi \|U^\phi - W\|_\square,$$

and

$$\delta_1(U,W) = \inf_\phi \|U^\phi - W\|_1$$

where ϕ ranges over all measure preserving bijections $[0,1] \to [0,1]$.

Both δ_1 and $\delta_\square$ are symmetric ($\delta_\square(U,W) = \delta_\square(W,U)$), and they satisfy the triangle inequality. However, they are only semimetrics, because different functions can have distance 0. To get a metric space, we have to identify pairs of functions at distance 0. We'll not carry this out formally.

To justify the above notation, we note that

$$\delta_\square(G_1,G_2) = \delta_\square(W_{G_1},W_{G_2}).$$

An argument similar to the proof of Theorem 4.1 gives that the sequence (G_n) of graphs converges to W if and only if $\delta_\square(W_{G_n},W) \to 0$ as $n \to \infty$. Furthermore, if W is a limit of the convergent sequence (G_n), then so is every function U with $\delta_\square(U,W) = 0$.

The $\delta_\square$ distance can be used to bound differences in subgraph densities, generalizing Lemma 3.3:

Lemma 4.4 *Let $U,W : [0,1]^2 \to [0,1]$ be two symmetric integrable functions. Then for every simple finite graph F,*

$$|t(F,U) - t(F,W)| \le |E(F)| \cdot \delta_\square(U,W).$$

Proof. Since the left hand side is invariant under measure-preserving transformations, it suffices to prove the seemingly weaker inequality

$$|t(F,U) - t(F,W)| \le |E(F)| \cdot \|U - W\|_\square.$$

Let $V(F) = [n]$ and $E(F) = \{i_1j_1,\dots,i_mj_m\}$. Then

$$t(F,U) - t(F,W) = \int_{[0,1]^n} \Big(\prod_{t=1}^m W(x_{i_t},x_{j_t}) - \prod_{i=1}^m U(x_{i_t},x_{j_t})\Big)\,dx.$$

We can write the integrand as

$$\sum_{s=0}^{m-1} \Big(\prod_{t\le s-1} W(x_{i_t},x_{j_t})\Big)\Big(\prod_{t>s} U(x_{i_t},x_{j_t})\Big)\times \times (W(x_{i_s},x_{j_s}) - U(x_{i_s},x_{j_s})).$$

To estimate the integral of a given term, let us fix all variables except x_{i_s} and x_{j_s}; then the integral is of the form

$$\int_{[0,1]^2} f(x_{i_s})g(x_{j_s})(W(x_{i_s},x_{j_s}) - U(x_{i_s},x_{j_s})),$$

where $0 \le f,g \le 1$. which is at most $\|U-W\|_\square$. Hence the whole integral of the term is at most $\|U-W\|_\square$, and so

$$|t(F,U) - t(F,W)| \le m\|U-W\|_\square.$$

□

Using this lemma and an argument similar to the proof of Theorem 4.1, we get the following convergence criterion:

Theorem 4.5 *Let $W \in \mathcal{W}_0$ and $W_n \in \mathcal{W}_0$ $(n = 1,2,\dots)$. Then $\delta_\square(W_n,W) \to 0$ if and only if $t(F,W_n) \to t(F,W)$ for every simple graph F.*

This theorem contains a certain converse of Lemma 4.4: it says that if two functions have approximately the same subgraph densities $t(F,W)$, then they are almost equal up to a measure-preserving transformation.

4.3.3 Compactness

Using the (weak) Regularity Lemma 3.1, one can prove [17] that

Theorem 4.6 *The metric space $(\mathcal{W}_0,\delta_\square)$ (where functions at distance 0 are identified) is compact.*

It turns out that stronger forms of the regularity lemma can in turn be derived from this compactness theorem.

4.3.4 Connections between the two distances

The distance $\delta_\square$ is quite different from the distance δ_1. Let $G_n = \mathbf{G}(n,1/2)$ be a random graph on n nodes with density 1/2. As we have seen in Example 5, this sequence is convergent with probability 1, and its limit is the function $W \equiv 1/2$. This means that

$$\delta_\square(W_{G_n},W) = \|W_{G_n} - W\|_\square \to 0$$

(since the function W is constant, measure-preserving transformations don't change $\|W_{G_n} - W\|_\square$). On the other hand, $W_{G_n} - W$ has value $\pm 1/2$ everywhere, and hence $\delta_1(W_{G_n},W) = \|W_{G_n} - W\|_1 = 1/2$. It is easy to see that no subsequence of this sequence has a limit in the δ_1 metric, which also implies that Theorem 4.6 does not hold for the δ_1 distance.

But there are some more subtle relations between the δ_1 and $\delta_\square$ distances. Inequality (3) implies that $\delta_\square$ is a continuous function with respect to the metric δ_1. It is much less trivial that the distance δ_1 is lower-semi-continuous with respect to the metric $\delta_\square$ [17]:

Lemma 4.7 *Suppose that both $\delta_\square(W_n,W) \to 0$ and $\delta_\square(U_n,U) \to 0$ as $n \to \infty$ (where $U,W,U_n,W_n \in \mathcal{W}_0$). Then $\delta_1(U,W) \le \liminf_n \delta_1(U_n,W_n)$.*

Proof. Let $\varepsilon > 0$. There exist stepfunctions Y and Z such that $\|U-Y\|_1 \le \varepsilon$ and $\|W-Z\|_1 \le \varepsilon$. Let Y and Z have at most k steps.

We may assume (by applying a measure-preserving transformation to U_n and W_n) that $\delta_\square(U_n,U) \le \|U_n - U\|_\square + \varepsilon/k^4$ and $\delta_\square(W_n,W) \le \|W_n - W\|_\square + \varepsilon/k^4$ for all n. If m is large enough, then $\|W_m - W\|_\square, \|U_m - U\|_\square \le \varepsilon/k^4$, and furthermore there exists a measure preserving transformation $\pi : [0,1] \mapsto [0,1]$ such that $\|U_m - W_m^\pi\|_1 \le \delta_1(U_m,W_m) + \varepsilon$. There is a partition $\mathcal{P}$ of $[0,1]$ into at most k^2 classes so that both stepfunctions Y and Z^π are constant on $S \times T$ for all

$S, T \in \mathcal{P}$. Then

$$\begin{aligned}
\|Y - Z^\pi\|_1 &= \sum_{S,T\in\mathcal{P}} \Big|\int_{S\times T}(Y - Z^\pi)\Big| \\
&\le \sum_{S,T\in\mathcal{P}} \Big|\int_{S\times T}(Y - U)\Big| + \sum_{S,T\in\mathcal{P}} \Big|\int_{S\times T}(U - U_m)\Big| \\
&\quad + \sum_{S,T\in\mathcal{P}} \Big|\int_{S\times T}(U_m - W_m^\pi)\Big| \\
&\quad + \sum_{S,T\in\mathcal{P}} \Big|\int_{S\times T}(W_m^\pi - W^\pi)\Big| \\
&\quad + \sum_{S,T\in\mathcal{P}} \Big|\int_{S\times T}(W^\pi - Z^\pi)\Big| \\
&\le \|Y - U\|_1 + k^4\|U - U_m\|_\square + \|U_m - W_m^\pi\|_1 \\
&\quad + k^4\|W_m^\pi - W^\pi\|_\square + \|W^\pi - Z^\pi\|_1 \\
&\le \delta_1(U_m, W_m) + 7\varepsilon.
\end{aligned}$$

Hence

$$\begin{aligned}
\delta_1(U, W) &\le \|U - W^\pi\|_1 \\
&\le \|U - Y\|_1 + \|Y - Z^\pi\|_1 + \|Z^\pi - W^\pi\|_1 \\
&\le \delta_1(U_m, W_m) + 9\varepsilon.
\end{aligned}$$

Since ε is arbitrary, this completes the proof. □

Finally, we remark that convergence in the $\|.\|_\square$ implies weak convergence in the following sense:

Lemma 4.8 *Let $W, W_1, W_2 \cdots \in \mathcal{W}_0$, and suppose that $\|W_n - W\|_\square \to 0$ as $n \to \infty$. Then for every integrable function $Z : [0,1]^2 \to \mathbb{R}$, we have*

$$\int_{[0,1]^2} Z(x,y)W_n(x,y)\,dx\,dy \to \int_{[0,1]^2} Z(x,y)W(x,y)\,dx\,dy.$$

In particular,

$$\int_S W_n \to \int_S W$$

for every measurable set $S \subseteq [0,1]^2$.

Proof. If Z is the indicator function of a rectangle, this follows from the definition of the $\|.\|_\square$ norm. Hence the conclusion follows for stepfunctions, since they are linear combinations of a finite number of indicator functions of rectangles. Then it follows for all integrable functions, since they are approximable in $L_1([0,1]^2)$ by stepfunctions. □

5. CHARACTERIZATIONS OF TESTABLE PARAMETERS

The following theorem [9] gives a number of equivalent conditions characterizing testability of a graph parameter.

Theorem 5.1 *For a simple graph parameter f, $0 \le f \le 1$, the following are equivalent:*

(a) *f is testable.*

(b) *For every $\varepsilon > 0$ there is an integer k_0 such that for every $k > k_0$ and every graph G on at least k nodes, a random set X of k nodes of G satisfies*

$$|f(G) - \mathsf{E}(f(G[X]))| < \varepsilon. \tag{4}$$

(c) *For every convergent graph sequence (G_n), the sequence of numbers $(f(G_n))$ is convergent.*

(d) *There exists a functional $\widehat{f}(W)$ on $\mathcal{W}_0$ that is continuous in the $\delta_\square$ distance of functions, and extends f in the sense that $|\widehat{f}(W_G) - f(G)| \to 0$ if $|V(G)| \to \infty$.*

(e) *For every $\varepsilon > 0$ there is an $\varepsilon_0 > 0$ and a positive integer n_0 so that if G_1 and G_2 are two graphs with $|V(G_i)| \ge n_0$ and $\delta_\square(G_1, G_2) < \varepsilon_0$, then $|f(G_1) - f(G_2)| < \varepsilon$.*

If we want to use (e) to prove that a certain parameter is testable, then the complicated definition of the $\delta_\square$ distance may cause a difficulty. So it is useful to show that (e) can be replaced by a weaker condition, which consists of three special cases of (e):

Supplement 5.2 *The following three conditions together are also equivalent to the testability of f:*

(e.1) *For every $\varepsilon > 0$ there is an $\varepsilon' > 0$ such that if G and G' are two simple graphs on the same node set and $d_\square(G, G') \le \varepsilon'$ then $|f(G) - f(G')| < \varepsilon$.*

(e.2) *Let $G(m)$ denote the graph obtained from G by blowing up each node into m twins. Then for every simple graph G, $f(G(m))$ has a limit as $m \to \infty$.*

(e.3) *Let G^+ be obtained from G by adding a single isolated node. Then $f(G^+) - f(G) \to 0$ if $|V(G)| \to \infty$.*

Proof Sketch. (a)⇒(b): The definition of testability is very similar to condition (b): it says, in this language, that a random set X of k nodes of G as in (b) satisfies

$$|f(G) - f(G[X])| < \varepsilon$$

with large probability. This clearly implies that this difference is small on the average.

(b)⇒(c): Suppose that a sequence (G_n) is convergent, then for any fixed k and sufficiently large n and m, and random subsets $X \subseteq V(G_n)$ and $Y \subseteq V(G_m)$ with $|X| = |Y| = k$, the distribution of $G_n[X]$ is very close to the distribution of $G_m[Y]$. Hence $|\mathsf{E}(f(G_n[X]) - \mathsf{E}(f(G_m[Y])| \le \varepsilon/3$. By (b), we can choose k large enough so that $|f(G_n) - \mathsf{E}(f(G_n[X]))| \le \varepsilon/3$ and $|f(G_m) - \mathsf{E}(f(G_m[Y]))| \le \varepsilon/3$, and so $|f(G_n) - f(G_m)| \le \varepsilon$.

(c)⇒(a): If (a) fails to hold, then there exists an $\varepsilon > 0$ such that for every k there exists a graph G_k on at least k nodes for which $|f(G_k) - f(G_k[X])| \ge \varepsilon$ holds with probability at least ε (where X is a random k-subset of $V(G_k)$). By Theorem 4.6, we may assume that G_k is convergent, and by Theorem 3.5 we know that if k is large enough, then $\delta_\square(G_k, G_k[X]) \le c(\log k)^{-1/3}$ with probability at least $1 - \varepsilon/2$. Hence we can fix a k-subset $X_k \subseteq V(G_k)$ such that both

$$|f(G_k) - f(G_k[X_k])| \ge \varepsilon \tag{5}$$

and

$$\delta_\square(G_k, G_k[X_k]) \le c(\log k)^{-1/3} \tag{6}$$

hold. Now merging the sequences (G_k) and $(G_k[X_k])$, we get a convergent graph sequence by (6), which violates (c) by (5).

(c)⇒(d): Consider any $W \in \mathcal{W}_0$, and define $\widehat{f}(W)$ as the limit of $f(G_n)$, where (G_n) is any sequence of graphs

converging to W. It is not hard to check that this value does not depend on the choice of the graph sequence, and that it satisfies the conditions.

(d)⇒(c): Consider a convergent graph sequence (G_n), and let $W \in \mathcal{W}$ be its limit. Then $\delta_\square(W_{G_n}, W) \to 0$ and so by the continuity of $\widehat{f}$, we have $\widehat{f}(W_{G_n}) - \widehat{f}(W) \to 0$. By assumption, $f(G_n) - \widehat{f}(W_{G_n}) \to 0$, and so $f(G_n) - \widehat{f}(W) \to 0$. This proves that $(f(G_n))$ is convergent.

(c)⇒(e): Suppose (e) does not hold, then there are two sequences (G_n) and (G'_n) of graphs so that $\delta_\square(G_n, G'_n) \to 0$ but $|f(G_n) - f(G'_n)| > c$ for some $c > 0$. By selecting a subsequence, we may assume that (G_n) is convergent. Then Theorem 4.1 implies that the merged sequence $(G_1, G'_1, G_2, G'_2 \dots)$ is convergent. But then the assumption $|f(G_n) - f(G'_n)| > c$ contradicts (c).

As we have remarked before, (e) implies all three of (e.1), (e.2) and (e.3). We conclude with sketching the proof that (e.1), (e.2) and (e.3) imply (c). Suppose not, then there exist two graph sequences (G_n) and (G'_n) such that $G_n, G'_n \to W$, but $|f(G_n) - f(G'_n)| > c$ for some fixed $c > 0$.

By (e.1), there exists an $\varepsilon > 0$ such that if G and G' are two graphs on the same node set, and $d_\square(G, G') \le \varepsilon$, then $|f(G) - f(G')| \le c/4$. By Corollary 3.6, for every n there is a simple graph H_n whose number of nodes k depends only on ε such that $\delta_\square(G_n, H_n) \le \varepsilon^4/(3C)$ (where C is the constant in (1)). By selecting an appropriate subsequence, we may assume that $H_n = H$ is the same graph for all n. Since (G_n) and (G'_n) have the same limit, it follows that $\delta_\square(G'_n, H) \le \varepsilon^4/(2C)$ for all n that are large enough.

Let us add to each G_n at most $k-1$ isolated nodes so that the resulting graph G^*_n has km_n nodes for some integer m_n. For n large enough, the m_n-fold blow-up $H[m_n]$ of H then satisfies

$$\delta_\square(G^*_n, H[m_n]) \le \varepsilon^4/C,$$

and so by (1), for a suitable overlay of G_n and $H[m_n]$, we have

$$d_\square(G^*_n, H[m_n]) \le \varepsilon,$$

and so, by the definition of ε,

$$|f(G^*_n) - f(H[m_n])| \le \frac{c}{4}.$$

Using (e.3), we see that $f(G^*_n) - f(G_n) \to 0$, and hence

$$|f(G_n) - f(H[m_n])| \le \frac{c}{3}$$

if n is large enough. Similarly, H has a m'_n-node blow-up $H[m'_n]$ such that

$$|f(G'_n) - f(H[m'_n])| \le \frac{c}{3}.$$

But since $H[m_n]$ and $H[m'_n]$ are blow-ups of the same graph H, (e.2) implies that $f(H[m_n]) - f(H[m'_n]) \to 0$, a contradiction. □

To illustrate the use of this theorem, let us consider the density $\mathsf{maxcut}(G)$ of the maximum cut. Of the conditions above, (b) and (c) do not seem easier to verify than the definition of testability (a). Conditions (e.1–3), on the other hand, are easy to verify: (e.1) is immediate from the definition of the $d_\square$ distance, and (e.3) is trivial. To prove (e.2), notice that every cut in G yields a cut in $G(m)$ with the same density, so $\mathsf{maxcut}(G(m)) \ge \mathsf{maxcut}(G)$. Conversely, consider a maximum cut C in $G(m)$. For each node $i \in V(G)$, select a node i' from the set of m twins in $G(m)$ corresponding to i. The nodes i' induce a copy of G, in which the cut C induces a cut C'. It is easy to check that the expected density of C' is the density of C. Hence $\mathsf{maxcut}(G(m)) = \mathsf{maxcut}(G)$.

We could also use (d): the parameter maxcut extends to functions in a natural way, so that the conditions in (d) are easily verified.

However, we don't get a polynomial bound (in $1/\varepsilon^2$) on the sample size as obtained in [2]. It is not clear which testable parameters can be tested by a polynomial sample size.

6. HEREDITARY PROPERTIES

Let $\mathcal{H}$ be a hereditary graph property, i.e., a property closed under removal of vertices. Let $d_{\text{edit}}(G, \mathcal{H})$ denote the "edit distance" of the graph G from property $\mathcal{H}$, i.e., the minimum number of edges to be changed in G (deleted and/or added) to get a graph with property $\mathcal{H}$. We need the normalized version $d_1(G, \mathcal{H}) = d_{\text{edit}}(G, \mathcal{H})/|V(G)|^2$. The following surprisingly general result was proved by Alon and Shapira [6].

Theorem 6.1 (Alon and Shapira) *For every hereditary property $\mathcal{H}$, the graph parameter $d_1(G, \mathcal{H})$ is testable.*

Our goal here is to sketch the proof of this theorem based on our analytic framework. One should point out that our proof is not constructive, in contrast to the proof of Alon and Shapira, which is constructive (but gives a "tower of towers" dependence of the test size on the error, through the use of a strengthened version of the Szemerédi Regularity Lemma).

6.1 The closure of a hereditary property

Let $\overline{\mathcal{H}}$ denote the set of functions in $\mathcal{W}_0$ for which there is a sequence (G_n) of graphs in $\mathcal{H}$ such that $G_n \to W$. It follows easily from Theorem 4.1 that $\overline{\mathcal{H}}$ is closed in the $\delta_\square$ distance. This implies that $\overline{\mathcal{H}}$ is invariant under measure-preserving transformations.

Lemma 6.2 *A function $W \in \mathcal{W}_0$ is in $\overline{\mathcal{H}}$ if and only if the W-random graph $\mathbf{G}(n, W) \in \mathcal{H}$ with probability 1 for every n.*

Proof. First, let $W \in \overline{\mathcal{H}}$, then there exists a sequence $G_n \in \mathcal{H}$ such that $G_n \to W$. Suppose that there is an n such that with positive probability, $\mathbf{G}(n, W)$ does not have property $\mathcal{H}$. Then there is a simple graph $F \notin \mathcal{H}$ such that $\mathbf{G}(n, W) = F$ with positive probability. On the other hand, we have $t_{\text{ind}}(F, G_n) = 0$ (since $\mathcal{H}$ is hereditary), which implies that $t_{\text{ind}}(F, W) = 0$, which means that the probability that $\mathbf{G}(n, W) = F$ is null.

Conversely, assume that $G(n, W)$ has property $\mathcal{H}$ with probability 1. Since $G(n, W) \to W$ with probability 1, it follows that $W \in \overline{\mathcal{H}}$. □

Corollary 6.3 *If $U \in \overline{\mathcal{H}}$, and $U' \in \mathcal{W}_0$ is a function such that $U'(x, y) = U(x, y)$ whenever $U(x, y) \in \{0, 1\}$, then $U' \in \overline{\mathcal{H}}$.*

6.2 Conclusion of the proof

To show that $d_1(.,\mathcal{H})$ is testable, we verify condition (d) in Theorem 5.1. The "extension to the limit" of $d_1(.,\mathcal{H})$ is the functional

$$d_1(W,\overline{\mathcal{H}}) = \inf_{U\in\overline{\mathcal{H}}} \|U-W\|_1,$$

where $\|U\|_1$ denotes the L_1 norm of the function U. Note that since $\overline{\mathcal{H}}$ is invariant under measure-preserving transformations, we could as well define this extension by

$$d_1(W,\overline{\mathcal{H}}) = \inf_{U\in\overline{\mathcal{H}}} \delta_1(U,W).$$

To verify that this is a valid choice for (d), it suffices to prove the following two claims.

Claim 6.4 *$d_1(W,\overline{\mathcal{H}})$ is a continuous function of W in the $\|.\|_\square$ norm.*

PROOF. We have to show that if $\|W-W_n\|_\square \to 0$, then

$$\lim_{n\to\infty} d_1(W_n,\overline{\mathcal{H}}) = d_1(W,\overline{\mathcal{H}}).$$

We may assume (by choosing a subsequence) that the limit on the left exists.

Let $\epsilon>0$ be an arbitrary number and let $U_n\in\overline{\mathcal{H}}$ be a sequence with $\|W_n-U_n\|_1 \le d_1(W_n,\overline{\mathcal{H}})+\epsilon$. We can assume by Theorem 4.6 that U_n converges to some function $U\in\mathcal{W}_0$ in $\delta_\square$. Then $U\in\overline{\mathcal{H}}$. Let π_n be a sequence of measure preserving transformations such that $\|U-U_n^{\pi_n}\|_\square\to 0$. Clearly $\delta_1(U_n^{\pi_n},W_n)\le\|U_n-W_n\|_1\le d_1(W_n,\overline{\mathcal{H}})+\epsilon$. It follows by Lemma 4.7 that $d_1(W,\overline{\mathcal{H}})\le\delta_1(U,W)\le\lim_n d_1(W_n,\overline{\mathcal{H}})+\epsilon$. This implies that $d_1(W,\overline{\mathcal{H}})\le\lim_n d_1(W_n,\overline{\mathcal{H}})$.

On the other hand, let $U\in\overline{\mathcal{H}}$ be function with $\|U-W\|_1\le d_1(W,\overline{\mathcal{H}})+\epsilon$. Let $S_0=U^{-1}(0)$ and $S_1=U^{-1}(1)$. Let Z_n be a function which is 0 on S_0, 1 on S_1 and is equal to W_n everywhere else. It follows by Corollary 6.3 that $Z_n\in\overline{\mathcal{H}}$. We have $d_1(W_n,\overline{\mathcal{H}})\le\|W_n-Z_n\|_1=\int_{S_0}W_n+\int_{S_1}(1-W_n)$. By Lemma 4.8, the right hand side converges to

$$\int_{S_0} W+\int_{S_1}(1-W)\le\|W-U\|_1\le d_1(W,\overline{\mathcal{H}})+\epsilon.$$

Hence $\lim_n d_1(W_n,\overline{\mathcal{H}})\le d_1(W,\overline{\mathcal{H}})$. □

Claim 6.5 *If $|V(G)|\to\infty$, then $d_1(G,\mathcal{H})\le d_1(W_G,\overline{\mathcal{H}})\le d_1(G,\mathcal{H})+o(1)$.*

PROOF. First we show that for every graph G,

$$d_1(G,\mathcal{H})\le d_1(W_G,\overline{\mathcal{H}}). \tag{7}$$

Let $\varepsilon>0$, and let $U\in\overline{\mathcal{H}}$ be such that $\|W_G-U\|_1\le d_1(W_G,\overline{\mathcal{H}})+\varepsilon$. By Corollary 6.3, we may assume that U is a $\{0,1\}$-valued function. Let $V(G)=[n]$, and let X_i be a uniform random element of the interval $L_i=[\frac{i-1}{n},\frac{i}{n}]$. Let G_X denote the graph on $[n]$ in which i and j are adjacent if and only if $U(X_i,X_j)=1$. Then with probability 1, G_X has property $\mathcal{H}$. Furthermore,

$$\begin{aligned}\mathsf{E}(d_1(G,G_X)) &= \frac{1}{n^2}\mathsf{E}(|E(G)\triangle E(G_X)|)\\ &= \frac{1}{n^2}\sum_{i,j=1}^n \Pr(W_G(X_i,X_j)\ne U(X_i,X_j))\\ &= \sum_{i,j=1}^n\int_{L_i\times L_j}|W_G(X_i,X_j)-U(X_i,X_j)|\\ &\le \|W_G-U\|_1\le d_1(W_G,\overline{\mathcal{H}})+\varepsilon.\end{aligned}$$

Hence there is a choice of X for which $G_X\in\mathcal{H}$ and $d_1(G,G_X)\le d_1(W_G,\overline{PP})+\varepsilon$. This proves (7).

To prove the other inequality, suppose that there exist a $c>0$ and a sequence of graphs (G_n) with $|V(G_n)|\to\infty$ such that $d_1(W_{G_n},\overline{\mathcal{H}})>d_1(G_n,\mathcal{H})+c$ for all n. Let H_n be a graph on $V(G_n)$ such that $H_n\in\mathcal{H}$ and $d_1(G_n,H_n)=d_1(G_n,\mathcal{H})$. We may assume (by selecting a subsequence) that $G_n\to W$ and $H_n\to U$ for some $U,W\in\mathcal{W}_0$. It follows by the definition of $\overline{\mathcal{H}}$ that $U\in\overline{\mathcal{H}}$. Furthermore, $\delta_\square(W_{G_n},W)\to 0$ and $\delta_\square(W_{H_n},U)\to 0$, hence by Lemma 4.7, we get that

$$\begin{aligned}d_1(W,\overline{\mathcal{H}})&\le\delta_1(W,U)\le\liminf_n\delta_1(W_{G_n},W_{H_n})\\ &\le\liminf_n d_1(G_n,H_n)=\liminf_n d_1(G_n,\mathcal{H}),\end{aligned}$$

a contradiction. □

Acknowledgement

We are grateful to Christian Avart, Vojtech Rödl and Mathias Schacht for pointing out an error in an earlier version of our proof of Theorem 6.1.

7. REFERENCES

[1] N. Alon, E. Fisher, M. Krivelevich and M. Szegedy: Efficient testing of large graphs, *Combinatorica* **20** (2000) 451–476.

[2] N. Alon, W. Fernandez de la Vega, R. Kannan and M. Karpinski: Random sampling and approximation of MAX-CSPs, *J. Comput. System Sci.* **67** (2003), 212–243.

[3] N. Alon and M. Krivelevich: Testing k-colorability, *SIAM J. Discrete Math. 15* (2002), 211–227.

[4] N. Alon, A. Naor: Approximating the Cut-Norm via Grothendieck's Inequality (preprint).

[5] N. Alon and A. Shapira, Every monotone graph property is testable, Proc. of the 37 ACM STOC, Baltimore, ACM Press (2005), to appear.

[6] N. Alon and A. Shapira: A Characterization of the (natural) Graph Properties Testable with One-Sided Error, Proc. FOCS 2005, to appear.

[7] S. Arora, D. Karger and M. Karpinski: Polynomial time approximation schemes for dense instnces of NP-hard problems, *Proc. 27-th ACM STOC* (1995), 284–293.

[8] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi: Counting graph homomorphisms (preprint).

[9] C. Borgs, J. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi: Convergent sequences of dense graphs, Parts I and II (manuscripts).

[10] F. Chung, R.L. Graham and R.M. Wilson: Quasi-random graphs, *Combinatorica* **9** (1989), 345–362.

[11] E. Fischer: The art of uninformed decisions: A primer to property testing, The Computational Complexity Column of the *Bulletin of the European Association for Theoretical Computer Science* **75** (2001), 97-126.

[12] A. Frieze and R. Kannan: Quick approximation to matrices and applications, *Combinatorica* **19**, 175–220.

[13] O. Goldreich, S. Goldwasser and D. Ron: Property testing and its connection to learning and approximation, *J. ACM* **45** (1998), 653–750.

[14] O. Goldreich and L. Trevisan: Three theorems regarding testing graph properties, *Random Structures and Algorithms*, **23** (2003), 23–57.

[15] L. Lovász and B. Szegedy: Limits of dense graph sequences (MSR Tech Report # MSR-TR-2004-79) ftp://ftp.research.microsoft.com/pub/tr/TR-2004-79.pdf

[16] L. Lovász and B. Szegedy: Graph limits and testing hereditary graph properties, Microsoft Research Technical Report MSR-TR-2005-110, ftp://ftp.research.microsoft.com/pub/tr/TR-2005-110.pdf

[17] L. Lovász and B. Szegedy: Szemerédi's lemma for the analist, http://research.microsoft.com/users/lovasz/analyst.pdf

[18] R. Rubinfeld and M. Sudan: Robust characterization of polynomials with applications to program testing, *SIAM J. on Computing* **25** (1996), 252–271.